Monika Sobejko

HOW CAN NATIVE SPEAKER CORPORA BE USED IN THE CLASSROOM IN A SKILLS-BASED APPROACH TO TEACHING ENGLISH?

Corpora are essentially electronically stored files of carefully selected samples of either spoken or written language, or a mixture of both. Using them to teach English is like going on a voyage, a voyage of discovery in which students play the role of 'language researchers'; this can be very motivating for them but time-consuming for teachers. Corpus data are relatively easy to access: they are literally a click of a mouse away and can frequently be used as raw data (print-outs of lines from a concordance can be brought to class or students can be asked to access a corpus and analyse the data on their own). However, teachers need to make a selection of what they want to present or have analysed, and they must also provide clear instructions or guiding questions for their students.

I would like to share a few tips that I hope will be useful to anyone who feels the need for supplementary materials for their course books. Language corpora are an excellent source of teaching materials and can be used to teach the four language skills as well as grammar and vocabulary (see O'Keeffe, McCarthy and Carter 2007). The article suggests several ways of using corpora to teach the skills of speaking, listening, reading and writing.

1. Using corpora to teach speaking and listening – chosen examples

Firstly, language corpora are a great source of authentic English and, unlike some of the invented course book dialogues, can help us bring 'real life' into the language classroom. We can give our students an opportunity to see how language is really used. I would like to focus on two examples. Consider this short exchange between a graduate student and a peer interviewer in the context of a research interview, taken (with minor editing) from the Michigan Corpus of Academic

Spoken English (MICASE, 2002).[1] I was looking for something appropriate for lower-intermediate students (levels A2-B1), so relatively short, and grammatically and lexically simple:

S1: […] *Were you working?*
S2: *I was looking for a job consulting and the market was pretty bad, so I didn't find anything, so I ended up working in a bookstore.*
S1: *Uhuh.*
S2: *Um, and which was kinda fun actually. It's something I've always wanted to do.*
S1: *Yeah.*
S2: *Um, it was neat to... to look at all the books as they come in.*
S1: *Oh yeah.*
S2: *And see the new ones.*
S1: *Oh yeah.*
S2: *And I get a pretty good discount on stuff so…*
S1: *Well, that's great.*

With this short sample of natural spoken English we can draw our students attention to the skill of 'active listenership', or 'the active, responsive role that listeners have in conversation'(O'Keeffe, McCarthy and Carter 2007: 142). O'Keeffe, McCarthy and Carter claim that 'good listenership' ought to be considered 'something which is both natural and desirable for efficient spoken communication' and suggest that, from a pedagogic perspective, there is a need to teach listenership as part of speaking and listening competence (O'Keeffe, McCarthy and Carter 2007: 142).

It is noticeable that in the dialogue S1 uses verbal signals to let S2 know that they understand what is being said and that their interlocutor is expected to continue rather than 'yield the floor' to S1 ('yeah', 'oh yeah', 'uhuh'). These verbal signals are referred to in studies in applied linguistics in a variety of ways: O'Keeffe, McCarthy and Carter refer to them as 'response tokens' (2007), other researchers introduce terms such as 'backchannel', 'minimal response' or 'listener response' (for more detailed discussion see O'Keeffe, McCarthy and Carter 2007) – what is important is their function in conversation, not word-class identity (i.e. whether they are adjectives, adverbs, etc).

We might claim that response tokens (verbal or non-verbal) are a natural way of reacting to our interlocutors in any language: we keep eye contact with them, nod, shrug, and so forth. But I would also argue that some response tokens might not be so easily transferrable from one language to another and, therefore, deserve more attention. Consider, for example, the English adverb *of course*, and its (apparent) Polish equivalent *oczywiście*. In Polish, it is often used to signal agreement to something or permission to do something (similar to *okay*, *yeah*, *no problem*

[1] MICASE can be accessed at: http://www.lsa.umich.edu/eli/micase/index.htm They also have an excellent section on teaching ESL/EAP materials and online sound files with accompanying transcripts. The website is well worth a visit.

in English). *Of course*, on the other hand, seems to be used somewhat differently in English, frequently meaning 'it goes without saying', 'it's self-evident', or 'obviously'.[2] As a result, native speakers of Polish have a tendency to overuse or misuse it, and in so doing might unintentionally sound rude or unfriendly towards their interlocutors. In such contexts it might be advisable to replace *of course* with *sure*, *right*, *yeah*, *okay* or *no problem*.

I have designed a simple matching activity for my lower-level students based on the sample dialogue from MICASE presented above, which asks them to fill in the gaps in the dialogue with S2's responses, as follows:

S1: […] *Were you working?*
S2:…….
S1*: Uhuh.*
S2:…….
S1*: Yeah.*
S2:……..
S2: *Oh yeah.*
S1:……..
S2: *Oh yeah.*
S1:……..
S2: *Well, that's great.*

S2's responses:

1. *Um, and which was kinda fun actually. It's something I've always wanted to do.*
2. *And see the new ones.*
3. *I was looking for a job consulting and the market was pretty bad, so I didn't find anything, so I ended up working in a bookstore.*
4. *And I get a pretty good discount on stuff so…*
5. *Um, it was neat to... to look at all the books as they come in.*

This matching exercise was followed up with a brief analysis of spoken language (vocabulary items such as "actually", "stuff", "neat", "pretty bad", "pretty good", "end up", examples of connected speech as in "kinda", hesitation, repetition on the part of S2), ways of avoiding repetition ("ones"), American versus British English ("bookstore", "neat"). I doubt whether any course book dialogue of comparable length would provide so many interesting characteristics to discuss. This kind of analysis can be followed by reading (and/or recording students), or asking students to continue the dialogue (for example, to write three or

[2] Compare, for example, the first definition of the meaning of *of course* in the (corpus-based, needless to say) Longman Dictionary of Contemporary English, which is as follows: 'used to show that other people probably already know what you are saying is true, or expect to hear it' (LDOCE, 4th ed., s.v. "of course"). Admittedly, at the moment I have no linguistic evidence of lack of pragmatic equivalence between the two adverbs and my observation is based only on my own intuition as a native speaker of Polish and that of an informant, who is a native speaker of British English. Personally, I think the problem of pragmatic equivalence of response tokens in Polish and English, or a comparative analysis of them, would constitute a fascinating area of study.

four more lines of the dialogue), or asking them to invent a few exchanges prior to the presented dialogue.

Here is another mini-example from MICASE, where we can ask students about appropriate intonation patterns and might practise reading/acting the dialogue in different ways depending on whether or not 'this' is a desirable item or whether S2 is really willing to share 'this' with S2:

S1: *can I have this?*
S2: *yeah*
S1: *yeah?*
S2: *yeah*
S1: *great*

Corpora are excellent sources of natural dialogues that can be exploited for a variety of purposes. I have found them particularly useful for teaching those discourse markers that are often altogether omitted or insufficiently practised in course books. A good example of such a marker is *like*.

Recently I was doing a listening comprehension activity with a group of upper-intermediate students. Having done the pre-listening and listening-for-gist activities with the students, we focused on the following snippet of a recording:

....and I'd read the Lostpedia entry about that episode, and I'm like I must've been half asleep, I didn't notice half the stuff that these super-experts, 'cuz they watch it ten times and they get all the details...[3]

Probably, owing to the fact that *like* occurred here in a rather unexpected context, as a discourse marker used to report the speaker's thoughts or feelings, the students had some difficulty picking it out. Therefore, as a pattern-noticing and awareness-raising exercise, I decided to show my students some more examples of *like* taken straight from MICASE:

I just, I look at him and [S1: mhm] he didn't have a beard then and I look at him I'm *like*[S1: <LAUGH>] did you go to Saint Lawrence? and [S1: uhuh] he's *like* yes and I'm *like*, oh and it just hit [S1: yeah] and, we're just so, *like* caught up in that we just completely ignored my advisor *like*, um, <S1: LAUGH> met once or twice, um Professor Whistler [S1: mhm] and, um, it was just really funny cuz we w- just sorta *like* catching up on, six years right there [S1: yeah] so so that was kinda kind of a surprise.

[italics mine]

The students were asked to read the above excerpt silently and find any differences as to the ways in which the word *like* was used in this exchange. Guided by my questions, they 'discovered' two ([1] and [3]) of the three rules concerning the use of *like* in spoken English as presented in Cambridge Grammar of English (Carter, McCarthy 2006: 101–102):

[3] Jimmy Wales, interview by Gareth Mitchell, Digital Planet, BBC [Originally aired November 11, 2009].

[1] *Like* is very commonly used (particularly among younger speakers) as a marker of reported speech, especially where the report involves a dramatic representation of someone's response or reaction:

So this bloke came up to me and I'm like 'Go away, I don't want to dance'.

[2] In some cases *like* acts as a 'filler', enabling the speaker to pause to think what to say next or to rephrase something. Pauses (...) can occur either side of the word:

They think that… like… by now we should be married and if we were married then it's okay like… to get on with your life and do what you want.

[3] *Like* can be placed in end position in order to qualify a preceding statement. It also indicates that the words chosen may not be appropriate:

Then she got out of the car all of a sudden like, and this bike hit her right in the back.

Next I asked the students to think of a recent conversation (with a friend, flatmate, teacher, etc) they could remember quite well and would be willing to report, and then, rather than using reported speech, they were supposed to 'report' their conversations in pairs in a 'more dramatic way', using *like* + direct speech (and their partners were to provide as much support as possible with 'uhuh', 'mhm', 'yeah', 'okay', etc.). From my observation of their performance I can say that they (1) had understood how to use the discourse marker very well, (2) really enjoyed the activity.

It seems that activities such as the above are very motivating for my students, and I hope will help them understand how spoken language really works.[4]

2. Using corpora to teach reading and writing

To fully exploit the potential of a language corpus to teach reading or writing, it would be best to build your own corpus. This can be done using specialist texts that your students will be particularly interested in, depending on their field of study, and then using this corpus to analyse individual lexical items, collocations, various cohesive devices, grammatical patterns occurring with greater frequency than usual (e.g. the passive voice in academic research papers), and the like. This not only makes the texts more motivating for students (as they directly concern their area of study or expertise), but also the results of frequency counts or concordances may be regarded as more reliable since students know exactly what type of texts are included, their length, level of difficulty, and lexical density (i.e. proportion of known to unknown words).[5]

[4] Having written that, though I must say that this inductive approach to teaching grammar or vocabulary for which language corpora seem to be ideally suited will not be equally well appreciated by every student – inevitably, some students will say: 'Why not just give us a rule? Why bother?'.

[5] For a more detailed analysis of the problem of building a language corpus and some practical suggestions I recommend taking the ELT Advantage online course (McCarthy et al. 2009).

An alternative is simply to access an existing corpus, and there are a number of these available online for non-commercial purposes (some free of charge, others for a subscription fee), as well as teacher-friendly tools designed to 'mine' such corpora.

One such tool is the VocabProfile program, which can be accessed at http://www.lextutor.ca/vp/eng. It analyses the vocabulary in a text and groups it into four categories: [1] the most frequently used 1,000 words in English; [2] the second most frequently used words in English (1,001–2,000); [3] words found in the Academic Word List[6] (this list consists of 570 families of words most frequently used in academic texts); and [4] remaining words (not found in categories 1–3). The typical ratios used by native speakers (in non-academic texts) are as follows: 70% (from 1); 10% (from 2); 10% (from 3); and 10% (from 4) (McCarthy et al. 2009 [online course]). It can be used to grade texts, depending on how difficult or how 'academic' they are. I use VocabProfile, for example, to help me decide which vocabulary to choose for supplementary exercises. McCarthy et al. (2009) suggest that students can be asked to have their own texts analysed by the program and compare their results to the above-mentioned typical native speaker ratios, and then use the analysis to decide which parts of their texts need to be rewritten, and finally rewrite those parts. Alternatively, texts can be analysed and then rewritten to 'fit' a specific audience (McCarthy et al. 2009 [online course]).

Another very useful tool is the Multi-Concordance program at the Compleat Lexical Tutor website, which can be found at http://www.lextutor.ca/concordancers/multi/. It allows access to a number of corpora (including Brown, British National Corpus (BNC) spoken, BNC written, Corpus_AWL, and graded corpora K1 and K2[7]) and, on the basis of the corpus data, generates gap-fills and quizzes. Simply go to the website and follow the instructions on the site. Here is an example of a short remedial exercise I prepared for my upper intermediate students with the help of Multi-Concordance (having realised that they still had some problems using the phrasal verbs listed below). The blanked out concordance lines are a classic example of what can quickly be prepared with the help of Multi-Concordance. Students must be instructed to read the lines top to bottom, looking carefully at what comes after and before the gap rather than read the lines from left to right as they are used to doing with a 'normal' text. – see Fig.1:

The above sets of concordance lines can also be used to present verbs: with a few guiding questions, students ought to be able to make informed guesses about the meaning and grammar of the verbs.

[6] The Academic Word List (AWL) can be accessed at http://www.victoria.ac.nz/lals/resources/ together with some suggestions concerning strategies of learning academic vocabulary using AWL. Students particularly interested in academic vocabulary can be directed to the website and encouraged to study independently of their classroom work.

[7] K1 – the most frequently used 1,000 words in English; K2 – the second most frequently used words in English (1,001–2,000)

Fig. 1.

Complete the following blanks with one of the phrasal verbs from the list below (there is one verb per one group of lines)

COME ACROSS, HOLD UP, GET BY, TAKE TO, KEEP UP WITH, FIT IN

with eating, our sleep patterns tend to have to _____with social expectations. Most people are
his adaptable nature he will find that he will _____well wherever he goes. The Horse pays a
to evaluate herself and work out where she _____with those around her. Doubtless you've had
they don't really belong. They need a girl who _____to make them feel that they do too

courtesy card they give to customers. `People _____it straight away," says David Morris,
a lovely bloke. I really like him. I really _____him straight away.
So I came as a library boy to Mr MX but I _____it really like a duck to water

We managed. We _____ We managed. We _____.
single mother by choice, who has enough money to _____ and supportive family and friends, can
the Kenilworth Road club, they will also have to _____without goalkeeper Alec Chamberlain, who
As the table shows, a family of four has to _____ on a little over £ 100 per week if

was walking through the jungle when he _____a poisonous snake. He walked cautiously round
hunting expedition in Hertfordshire, they _____a 19th-century chapel. Despite dust, damp and
their sister. Just the other day I _____them playing a game together. Nicholas had

Carla halted in a line of vehicles _____at the traffic lights on Golders Green
notice, and if you can't make it, or you get _____or something, I'll ring you just as soon as
of up to seven hours, and freight is being _____for twelve hours or more. FERRY STRIKE

at university and having to work fairly hard to _____the demands of her course. Their
the age when there is a feeling that `I cannot _____the younger ones and don't want to be
cut grass, water and weed, and generally try to _____events in the garden. It is a
at what happens when you do not or cannot _____the payments later. Even if the court

Key to exercise in fig. 1

COME ACROSS, HOLD UP, GET BY, TAKE TO, KEEP UP WITH, FIT IN

with eating, our sleep patterns tend to have to **fit in with** social expectations. Most people are
his adaptable nature he will find that he will **fit in** well wherever he goes. The Horse pays a
to evaluate herself and work out where she **fitted in with** those around her. Doubtless you've had
they don't really belong. They need a girl who **fits in** to make them feel that they do too

courtesy card they give to customers. "People **take to** it straight away," says David Morris,
a lovely bloke. I really like him. I really **took to** him straight away.
So I came as a library boy to Mr MX but I **took to** it really like a duck to water

We managed. We **got by** We managed. We **got by**.
single mother by choice, who has enough money to **get by** and supportive family and friends, can
the Kenilworth Road club, they will also have to **get by** without goalkeeper Alec Chamberlain, who
As the table shows, a family of four has to **get by on** a little over £ 100 per week if

was walking through the jungle when he **came across** a poisonous snake. He walked cautiously round
hunting expedition in Hertfordshire, they **came across** a 19th-century chapel. Despite dust, damp and
their sister. Just the other day I **came across** them playing a game together. Nicholas had

Carla halted in a line of vehicles **held up** at the traffic lights on Golders Green
notice, and if you can't make it, or you get **held up** or something, I'll ring you just as soon as
of up to seven hours, and freight is being **held up** for twelve hours or more. FERRY STRIKE

at university and having to work fairly hard to **keep up with** the demands of her course. Their
the age when there is a feeling that `I cannot **keep up with** the younger ones and don't want to be
cut grass, water and weed, and generally try to **keep up with** events in the garden. It is a
at what happens when you do not or cannot **keep up with** the payments later. Even if the court

Alternatively, the exercise can be made easier for weaker students, who can be asked to match the verbs with the definitions provided by the teacher. For example, the definitions for "keep up with" might look as follows:

1. to move, make progress or increase at the same rate as sb/sth.
2. to continue to pay or do sth regularly.
3. to learn about or be aware of the news, current events.

Students can then be asked to divide the verbs into four categories (intransitive phrasal verbs, inseparable transitive phrasal verbs with a particle, separable transitive phrasal verbs with a particle, phrasal verbs with a particle and preposition).

It is also possible to point out verb + adverb collocations ("take to something straight away", "fit in well") or idioms ("take to something like a duck to water").

My observations have led me to believe that students often find it difficult to use phrasal verbs correctly, and this exposure to multiple contexts (made possible with the help of concordance lines) can give them a better opportunity to observe how and in what contexts those verbs tend to be used. How much extensive reading would they have to do to find as many examples of use of these particular verbs? It is also a good way to practice reading (deducing meaning from co-text) and critical thinking skills.

On the Compleat Lexical Tutor site, there are two more useful programs: one is Range, which 'tells you about the distribution of words or other lexical units across a set of two or more texts. The texts can be comparable corpora or subdivisions of a corpus, or a set of texts supplied by a user' (Compleat Lexical Tutor. Range. http://www.lextutor.ca/range/ [accessed December 29, 2009]; the other is Phrase Extractor, which can be found under the link N-Gram on the same site. The latter is a very helpful tool if we want to look for chunks of words (2, 3, or 4-word chunks) in a text and prioritise them on the basis of their frequency.

Conclusions

These are simply a few suggestions as to how native speaker corpora, tools and resources that I have used myself and can therefore recommend, might be used in teaching language skills. My intention was to draw attention to these somewhat underrated but nonetheless very helpful, motivating and easy to use teaching resources, as well as to a few teacher-friendly online tools for working with corpora. Their number is growing and all we need to do is look out for them and add to our repertoire of teaching resources.

Listed below are links to websites containing materials which I have found interesting, and a short (by no means exhaustive) list of language corpora.

The University of Warwick: The Academic English Zone

http://www2.warwick.ac.uk/fac/soc/al/learning_english/activities/aez/resources

Literacynet.org: Learning Resources
http://literacynet.org/cnnsf/blackouts/home.html
The site offers web-based activities using CNN and CBS news stories.

Tim Johns' Data-Driven Learning page
http://www.ecml.at/projects/voll/our_resources/graz_2002/ddrivenlrning/index.htm
The site offers plenty of excellent DDL materials, reading material and links to other useful sites.

The Ohio University, Department of Linguistics:
http://www.ohio.edu/linguistics/esl/writing/index.html
The site offers a plethora of articles on teaching/learning language skills, including academic writing, giving further links to follow.

The Complete Lexical Tutor
http://www.lextutor.ca/
On this website, there are a number of software packages that interface with corpora, including software that allows you to load your own texts, do vocabulary tests based on vocabulary bands and levels, or access graded readers.

The Virtual Language Centre
http://vlc.polyu.edu.hk/concordance/WWWConcappE.htm
On this website, you can search several different corpora, including some learner corpora (samples of student writing), and generate concordances.

The Academic Word List
http://www.victoria.ac.nz/lals/resources/
The website provides access to the list of academic vocabulary developed by Averil Coxhead.

A short list of freely-available online language corpora:

The British National Corpus
http://www.natcorp.ox.ac.uk/
A 100 million word collection of samples of written and spoken language from a wide range of sources, designed to represent a wide cross-section of current British English, both spoken and written.

The Collins WordbanksOnline English corpus sampler
http://www.collins.co.uk/Corpus/CorpusSearch.aspx
Composed of 56 million words of contemporary written and spoken text. You can type in some simple queries here and get a display of concordance lines from the corpus.

The Corpus of Contemporary American English (COCA)
http://www.americancorpus.org/
Contains more than 400 million words of text and is equally divided among spoken, fiction, popular magazines, newspapers, and academic texts. It is a freely-available corpus.

Michigan Corpus of Academic Spoken English (MICASE)
http://www.lsa.umich.edu/eli/micase/index.htm
Consists of 1.8 million words of spoken academic English recorded at the University of Michigan. It includes lectures, labs, seminars, dissertation defenses, interviews, meetings, tutorials and service encounters. It is freely-available via the Web.

Bibliography

Carter R., McCarthy M., 2006. *Cambridge Grammar of English*. Cambridge: Cambridge University Press.

O'Keeffe A., McCarthy M., Carter R., 2007. *From corpus to classroom*. Cambridge: Cambridge University Press.

Simpson R.C., Briggs S.L., Ovens J., Swales J.M., 2002. *The Michigan Corpus of Academic Spoken English*. Ann Arbor, MI: The Regents of the University of Michigan.

McCarthy M., O'Keeffe A., Walsh S., 2009. *An Introduction to Corpora in English Language Teaching*. [ELT Advantage online course, accessed at http://www.ed2go.com/eltadvantage October, 21 – December, 12, 2009].

Wales J., interview by G. Mitchell, Digital Planet, BBC [Originally aired on November 11, 2009].